TITLE OF THE INVENTION

SPEECH SYNTHESIS APPARATUS, CONTROL METHOD THEREFOR, AND COMPUTER-READABLE MEMORY

BACKGROUND OF THE INVENTION

The present invention relates to a speech synthesis apparatus which has a database for managing phonemic piece data and performs speech synthesis by using the phonemic piece data managed by the database, a control method for the apparatus, and a computer-readable memory.

As a conventional speech synthesis method, a synthesis method based on a waveform concatenation scheme is available. In the waveform concatenation synthesis method, the prosody is changed by the pitch synchronous waveform overlap adding method of pasting waveform element pieces corresponding one to several pitches at desired pitch intervals. The waveform concatenation synthesis method can obtain more natural synthetic speech than a synthesis method based on a parametric scheme, but suffers the problem of a narrow allowable range with respect to changes in prosody.

Under the circumstances, attempts are made to improve the speech quality by preparing various speech data and properly selecting and using them. As a criterion for selection of speech data, information such as a phonemic context (a phoneme to be synthesized or a few phonemes on two sides of the target phoneme) or a fundamental frequency

F0 is used.

The following problems are, however, posed in the above conventional speech synthesis method.

If, for example, there is no data that satisfies a phonemic context as a synthesis target, a search for necessary speech data is made again by relaxing the condition associated with the phonemic context. The execution of this re-search in speech synthesis complicates the processing, resulting in an increase in processing time. In addition, when the fundamental frequency F0 is to be used as a criterion for selection of speech data, each speech data must be evaluated in association with the fundamental frequency F0 to obtain speech data that matches most with the fundamental frequency F0 of the speech data to be synthesized.

SUMMARY OF THE INVENTION

The present invention has been made in consideration of the above problems, and has as its object to provide a speech synthesis apparatus capable of performing speech synthesis with high precision at high speed, a control method therefor, and a computer-readable memory.

In order to achieve the above object, a speech synthesis apparatus according to the present invention has the following arrangement.

There is provided a speech synthesis apparatus having a database for managing phonemic piece data, comprising:

generating means for generating a second phoneme in consideration of a phonemic context for a first phoneme as a search target;

search means for searching the database for a phonemic piece data corresponding to the second phoneme;

re-search means for generating a third phoneme by changing the phonemic context on the basis of the search result obtained by the search means, and re-searching the database for phonemic piece data corresponding to the third phoneme; and

registration means for registering the search result obtained by the search means or the re-search means in a table in correspondence with the second or third phoneme.

In order to achieve the above object, a speech synthesis apparatus according the present invention has the following arrangement.

There is provided a speech synthesis apparatus for performing speech synthesis by using phonemic piece data managed by a database, comprising:

storage means for storing a table for managing position information indicating a position of phonemic piece data in the database in correspondence with a phoneme obtained in consideration of a phonemic context made to correspond to the phonemic piece data;

calculation means for acquiring each phonemic context information of a phoneme group as a synthesis target and

fundamental frequencies corresponding thereto and calculating an average of acquired fundamental frequencies;

search means for searching a phoneme group corresponding to the phonemic context information from the table;

acquisition means for acquiring, from the table, position information of phonemic piece data corresponding to a predetermined phoneme of the phoneme group searched out by the search means, on the basis of the average of fundamental frequencies calculated by the calculation means; and

changing means for acquiring phonemic piece data indicated by the position information acquired by the acquisition means from the database, and changing a prosody of the acquired phonemic piece data.

In order to achieve the above object, a control method for a speech synthesis apparatus according to the present invention has the following steps.

There is provided a control method for a speech synthesis apparatus having a database for managing phonemic piece data, comprising:

the generating step of generating a second phoneme in consideration of a phonemic context for a first phoneme as a search target;

the search step of searching the database for a phonemic piece data corresponding to the second phoneme;

the re-search step of generating a third phoneme by

changing the phonemic context on the basis of the search result obtained in the search step, and re-searching the database for phonemic piece data corresponding to the third phoneme; and

the registration step of registering the search result obtained in the search step or the re-search step in a table in correspondence with the second or third phoneme.

In order to achieve the above object, a control method for a speech synthesis apparatus according to the present invention has the following steps.

There is provided a control method for a speech synthesis apparatus for performing speech synthesis by using phonemic piece data managed by a database, comprising:

the storage step of storing a table for managing position information indicating a position of phonemic piece data in the database in correspondence with a phoneme obtained in consideration of a phonemic context made to correspond to the phonemic piece data;

the calculation step of acquiring each phonemic context information of a phoneme group as a synthesis target and fundamental frequencies corresponding thereto and calculating an average of acquired fundamental frequencies;

the search step of searching a phoneme group corresponding to the phonemic context information from the table;

the acquisition step of acquiring, from the table,

position information of phonemic piece data corresponding to a predetermined phoneme of the phoneme group searched out in the search step, on the basis of the average of fundamental frequencies calculated in the calculation step; and

the changing step of acquiring phonemic piece data indicated by the position information acquired in the acquisition step from the database, and changing a prosody of the acquired phonemic piece data.

In order to achieve the above object, a computer-readable memory according to the present invention has the following program codes.

There is provided a computer-readable memory storing program codes for controlling a speech synthesis apparatus having a database for managing phonemic piece data, comprising:

a program code for the generating step of generating a second phoneme in consideration of a phonemic context for a first phoneme as a search target;

a program code for the search step of searching the database for a phonemic piece data corresponding to the second phoneme;

a program code for the re-search step of generating a third phoneme by changing the phonemic context on the basis of the search result obtained in the search step, and re-searching the database for phonemic piece data corresponding to the third phoneme; and

a program code for the registration step of registering the search result obtained in the search step or the re-search step in a table in correspondence with the second or third phoneme.

In order to achieve the above object, a computer-readable memory according to the present invention has the following program codes.

There is provided a computer-readable memory storing program codes for controlling a speech synthesis apparatus for performing speech synthesis by using phonemic piece data managed by a database, comprising:

a program code for the storage step of storing a table for managing position information indicating a position of phonemic piece data in the database in correspondence with a phoneme obtained in consideration of a phonemic context made to correspond to the phonemic piece data;

a program code for the calculation step of acquiring each phonemic context information of a phoneme group as a synthesis target and fundamental frequencies corresponding thereto and calculating an average of acquired fundamental frequencies;

a program code for the search step of searching a phoneme group corresponding to the phonemic context information from the table;

a program code for the acquisition step of acquiring, from the table, position information of phonemic piece data

corresponding to a predetermined phoneme of the phoneme group searched out in the search step, on the basis of the average of fundamental frequencies calculated in the calculation step; and

a program code for the changing step of acquiring phonemic piece data indicated by the position information acquired in the acquisition step from the database, and changing a prosody of the acquired phonemic piece data.

According to the present invention described above, a speech synthesis apparatus capable of performing speech synthesis with high precision at high speed, a control method therefor, and a computer-readable memory can be provided.

Other features and advantages of the present invention will be apparent from the following description taken in conjunction with the accompanying drawings, in which like reference characters designate the same or similar parts throughout the figures thereof.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a block diagram showing the arrangement of a speech synthesis apparatus according to the first embodiment of the present invention; ,

Fig. 2 is a flow chart showing search processing executed in the first embodiment of the present invention;

Fig. 3 is a view showing an index managed in the first embodiment of the present invention;

Fig. 4 is a flow chart showing speech synthesis processing executed in the first embodiment of the present invention;

Fig. 5 is a view showing a table obtained from the index managed in the first embodiment of the present invention;

Fig. 6 is a flow chart showing search processing executed in the second embodiment of the present invention;

Fig. 7 is a view showing an index managed in the second embodiment of the present invention:

Fig. 8 is a flow chart showing search processing executed in the third embodiment of the present invention; and

Fig. 9 is a view showing an index managed in the third embodiment of the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Fig. 1 is a block diagram showing the arrangement of a speech synthesis apparatus according to the first embodiment of the present invention.

Reference numeral 103 denotes a CPU for performing numerical operation/control, control on the respective components of the apparatus, and the like, which are executed in the present invention; 102, a RAM serving as a work area for processing executed in the present invention and a temporary saving area for various data; 101, a ROM storing various control programs such as programs executed in the

present invention, and having an area for storing a database 101a for managing phonemic piece data used for speech synthesis; 109, an external storage unit serving as an area for storing processed data; and 105, a D/A converter for converting the digital speech data synthesized by the speech synthesis apparatus into analog speech data and outputting it from a loudspeaker 110.

Reference numeral 106 denotes a display control unit for controlling a display 111 when the processing state and processing results of the speech synthesis apparatus, and a user interface are to be displayed; 107, an input control unit for recognizing key information input from a keyboard 112 and executing the designated processing; 108, a communication control unit for controlling transmission/reception of data through a communication network 113; and 104, a bus for connecting the respective components of the speech synthesis apparatus to each other.

Search processing of searching for a target phoneme, of the processing executed in the first embodiment, will be described next with reference to Fig. 2.

Fig. 2 is a flow chart showing search processing executed in the first embodiment of the present invention.

In the first embodiment, as phonemic contexts, two phonemes on both sides of each phoneme, i.e., phonemes as right and left phonemic contexts called a triphone, are used.

First of all, in step S1, a phoneme p as a search target

from the database 101a is initialized to a triphone ptr. In step S2, a search is made for the phoneme p from the database 101a. More specifically, a search is made for phonemic piece data having label p indicating the phoneme p. It is then checked in step S4 whether there is the phoneme p in the database 101a. If it is determined that the phoneme p is not present (NO in step S4), the flow advances to step S3 to change the search target to a substitute phoneme having lower phonemic context dependency than the phoneme p. If the phoneme p matching with the triphone ptr is not present in the database 101a, the phoneme p is changed to the right phonemic context dependent phoneme. If the right phonemic context dependent phoneme does not match with the triphone ptr, the phoneme p is changed to the left phonemic context dependent phoneme. If the left phonemic context dependent phoneme does not match with the triphone ptr, the phoneme p is changed to another phoneme independently of a phonemic context. Alternatively, a high priority may be given to a left phonemic context phoneme for a vowel, and a high priority may be given to a right phonemic context phoneme for a consonant. In addition, if there is no phoneme p that matches with the triphone ptr, one or both of left and right phonemic contexts may be replaced with similar phonemic contexts. For example, the "k" (consonant of the "ka" column in the Japanese syllabary) may be used as a substitute when the right phonemic context is "p" (consonant for the "pa" column which is

modified "ha" column in the Japanese syllabary). Note, the Japanese syllabary is the Japanese basic phonetic character set. The character set can be arranged in a matrix where there are five (5) rows and ten (10) columns. The five rows are respectively the five vowels of the English language and the ten rows consist of 9 consonants and the column of the five vowels. A phonetic (sound) character is represented by the sound resulting from combining a column character and a row character, e.g. column "t" and row "e" is pronounced "te"; column "s" and row "o" is pronounced "so". After the phoneme p as the search condition is changed in this manner, the flow returns to step S2.

If it is determined that the phoneme p is present (YES in step S4), the flow advances to step S5 to calculate a mean F0 (the mean of the fundamental frequencies from the start of phonemic piece data to the end). Note that this calculation may be performed with respect to a logarithm F0 (function of time) or linear F0. Furthermore, the mean F0 of unvoiced speech may be set to 0 or estimated from the mean F0 of phonemic piece data of phonemes on both sides of the phoneme p by some method.

In step S6, the respective searched phonemic piece data are aligned (sorted) on the basis of the calculated mean F0. In step S7, the sorted phonemic piece data are registered in correspondence with the triphone ptr. As a result of registration, an index like the one shown in Fig. 3 is

obtained, which indicates the correspondence between generated phonemic piece data and triphones. As shown in Fig. 3, in the pointers managed in correspondence with the triphones, "phonemic piece position" indicating the location of each phonemic piece data in the database 101a and "mean F0" are managed in the form of a table.

Steps S1 to S7 are repeated for all conceivable triphones. It is then checked in step S8 whether the processing for all the triphones is complete. If it is determined that the processing is not complete (NO in step S8), the flow returns to step S1. If it is determined that the processing is complete (YES in step S8), the processing is terminated.

Speech synthesis processing of performing speech synthesis by searching for phonemic piece data of a phoneme as a synthesis target using the index generated by the processing described with reference to Fig. 2 will be described next with reference to Fig. 4.

Fig. 4 is a flow chart showing the speech synthesis processing executed in the first embodiment of the present invention.

When speech synthesis processing is to be performed, the triphone context ptr of the phoneme p as a synthesis target and F0 trajectory are given. Speech synthesis is then performed by searching phonemic piece data of phonemes on the basis of mean F0 and triphone context ptr and using the

waveform overlap adding method.

First of all, in step S9, mean F0' which is mean of the given F0 trajectory of a synthesis target is calculated. In step S10, a table for managing the phonemic piece position of phonemic piece data corresponding to the triphone ptr of the phoneme p is searched out from the index shown in Fig. 3. If, for example, the triphone ptr is "a. A. b", the table shown in Fig. 5 is obtained from the index shown in Fig. 3. Since proper substitute phonemes have been obtained by the above search processing, the result of this step never becomes empty.

In step S11, the phonemic piece position of phonemic piece data having the mean F0 nearest to the mean F0' is obtained on the basis of the table obtained in step S10. In this case, since the phonemic piece data have been sorted by the above search processing on the basis of mean F0, a search can be made by using a binary search method or the like. In step S12, phonemic piece data is retrieved from the database 101a in accordance with the phonemic piece position obtained in step S11. In step S13, the prosody of the phonemic piece data obtained in step S12 is changed by using the waveform overlap adding method.

As described above, according to the first embodiment, when the absence of phonemic piece data is determined after the presence/absence of phonemic piece data is checked with respect to all the conceivable phonemic contexts, the

processing is simplified and the processing speed is increased by preparing substitute phonemes in advance. In addition, since information associated with the mean F0 of phonemic piece data present in each phonemic context is extracted in advance, and the phonemic piece data are managed on the basis of the extracted information. This can increase the processing speed of speech synthesis processing.

[Second Embodiment]

Quantization of the mean F0 of phonemic piece data may replace calculation of the mean F0 of continuous phonemic piece data in step S5 in Fig. 2 in the first embodiment. This processing will be described with reference to Fig. 6.

Fig. 6 is a flow chart showing search processing executed in the second embodiment of the present invention.

Note that the same step numbers in Fig. 6 denote the same processes as those in Fig. 2 in the first embodiment, and a detailed description thereof will be omitted.

In step S14, a mean F0 of the phonemic piece data of searched phonemes p is quantized to obtain the quantized mean F0 (obtained by quantizing the mean F0 as a continuous value at certain intervals). This calculation may be performed for the logarithm F0 or linear F0. In addition, the mean F0 of unvoiced speech may be set to 0, or unvoiced speech may be estimated from the mean F0 of phonemic piece data on both side of the unvoiced speech by some method.

In step S6a, the searched phonemic piece data are

aligned (sorted) on the basis of the quantized mean F0. In step S7a, the sorted phonemic piece data are registered in correspondence with triphones ptr. As a result of registration, an index indicating the correspondence between the generated phonemic piece data and the triphones is formed as shown in Fig. 7. In addition, as shown in Fig. 7, in the pointers managed in correspondence with the triphones, "phonemic piece position" indicating the location of each phonemic piece data in the database 101a and "mean F0" are managed in the form of a table.

Steps S1 to S7a are repeated for all possible triphones. It is then checked in step S8a whether the processing for all the triphones is complete. If it is determined that the processing is not complete (NO in step S8a), the flow returns to step S1. If it is determined that the processing is complete (YES in step S8a), the processing is terminated.

As described above, according to the second embodiment, in addition to the effects obtained in the first embodiment, the number of phonemic pieces and the calculation amount for search processing can be reduced by using the quantized mean F0 of phonemic piece data.

[Third Embodiment]

In the second embodiment, after the portions between the sorted phonemic piece data are interpolated, the respective phonemic piece data may be registered in correspondence with the triphones ptr. That is, an

arrangement may be made such that phonemic piece positions corresponding to the quantized means F0 of all the quantized phonemic piece data can be searched out in the tables in the index. This processing will be described with reference to Fig. 8.

Fig. 8 is a flow chart showing search processing executed in the third embodiment of the present invention.

Note that the same step numbers in Fig. 8 denote the same processes as those in Fig. 6 in the second embodiment, and a detailed description thereof will be omitted.

In step S15, the portions between sorted phonemic piece data are interpolated. In step S7b, the interpolated phonemic piece data are registered in correspondence with triphones ptr. As a result of registration, an index indicating the correspondence between the generated phonemic piece data and the triphones is formed as shown in Fig. 9. In addition, as shown in Fig. 9, in the pointers managed in correspondence with the triphones, "phonemic piece position" indicating the location of each phonemic piece data in the database 101a and "mean F0" are managed in the form of a table.

Steps S1 to S7b are repeated for all possible triphones. It is then checked in step S8b whether the processing for all the triphones is complete. If it is determined that the processing is not complete (NO in step S8b), the flow returns to step S1. If it is determined that the processing is complete (YES in step S8b), the processing is terminated.

As described above, according to the third embodiment, in addition to the effects obtained in the second embodiment, since the phonemic piece positions of all phonemic piece data are managed, the processing in step S11 in Fig. 4 can be simply implemented as the step of referring to a table. This can further simplify the processing.

Note that the present invention may be applied to either a system constituted by a plurality of equipments (e.g., a host computer, an interface device, a reader, a printer, and the like), or an apparatus consisting of a single equipment (e.g., a copying machine, a facsimile apparatus, or the like).

The objects of the present invention are also achieved by supplying a storage medium, which records a program code of a software program that can realize the functions of the above-mentioned embodiments to the system or apparatus, and reading out and executing the program code stored in the storage medium by a computer (or a CPU or MPU) of the system or apparatus.

In this case, the program code itself read out from the storage medium realizes the functions of the above-mentioned embodiments, and the storage medium which stores the program code constitutes the present invention.

As the storage medium for supplying the program code, for example, a floppy disk, hard disk, optical disk, magneto-optical disk, CD-ROM, CD-R, magnetic tape,

nonvolatile memory card, ROM, and the like may be used.

The functions of the above-mentioned embodiments may be realized not only by executing the readout program code by the computer but also by some or all of actual processing operations executed by an OS (operating system) running on the computer on the basis of an instruction of the program code.

Furthermore, the functions of the above-mentioned embodiments may be realized by some or all of actual processing operations executed by a CPU or the like arranged in a function extension board or a function extension unit, which is inserted in or connected to the computer, after the program code read out from the storage medium is written in a memory of the extension board or unit.

As many apparently widely different embodiments of the present invention can be made without departing from the spirit and scope thereof, it is to be understood that the invention is not limited to the specific embodiments thereof except as defined in the appended claims.